

091533,255

ABSTRACT OF THE DISCLOSURE

A table in an HTML document is analyzed to generate cell position data indicating a positional relationship between cells and cell vectors

5      representing characteristics of the cells, and a table type is judged with reference to the cell position data and the cell vectors, and, if the table type is a table describing a table, it is judged whether the data is represented in a column or a row with

10     reference to the cell position data and the cell vectors, and a cut direction of the table is determined, and segments are generated with reference to the table type and the cut direction. If the table type is a table for layout, the cells are clustered

15     with reference to the cell vectors, and the segments are generated with reference to the cell position data and cell cluster information.